

AL-TP-1993-0011

AD-A264 021



2

ARMSTRONG

**PREDICTING AIRCREW TRAINING  
PERFORMANCE WITH PSYCHOMETRIC g**

Michele Morales Olea  
Malcolm James Ree



**HUMAN RESOURCES DIRECTORATE  
TECHNICAL TRAINING RESEARCH DIVISION  
7909 Lindbergh Drive  
Brooks Air Force Base, TX 78235-5352**

LABORATORY

April 1993

Interim Technical Paper for Period January 1992 - June 1992

Approved for public release; distribution is unlimited.

93 5 11 05 2

93-10291



**AIR FORCE MATERIEL COMMAND  
BROOKS AIR FORCE BASE, TEXAS**

## NOTICES

This technical paper is published as received and has not been edited by the technical editing staff of the Armstrong Laboratory.

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Office of Public Affairs has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

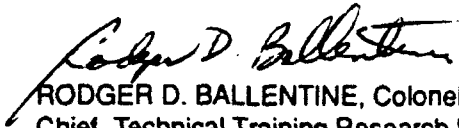
This paper has been reviewed and is approved for publication.



MICHELE MORALES OLEA  
Project Scientist



MARK S. TEACHOUT, Senior Scientist  
Skills Development Branch



RODGER D. BALLENTINE, Colonel, USAF  
Chief, Technical Training Research Division

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
<small>Public reporting burden for this report is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.</small>				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE April 1993		3. REPORT TYPE AND DATES COVERED Interim January 1992 - June 1992
4. TITLE AND SUBTITLE  Predicting Aircrew Training Performance with Psychometric <i>g</i>			5. FUNDING NUMBERS  PE - 62205F PR - 1121 TA - 12 WU - 00	
6. AUTHOR(S)  Michele Morales Olea Malcolm James Ree				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Armstrong Laboratory Human Resources Directorate Technical Training Research Division 7909 Lindbergh Drive Brooks Air Force Base, TX 78235-5352			8. PERFORMING ORGANIZATION REPORT NUMBER  AL-TP-1993-0011	
9. SPONSORING MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSORING MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION AVAILABILITY STATEMENT  Approved for public release; distribution is unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words)  A comparison of the validity of general cognitive ability, <i>g</i> , and specific ability, <i>s</i> , for predicting pilot and navigator training success revealed that <i>g</i> was the best predictor for all ten criteria and that <i>s</i> contributed little beyond <i>g</i> . The criteria included both academic performance, flying maneuvers, and airborne navigation.				
14. SUBJECT TERMS Cognitive ability Psychometric <i>g</i> Training principal components			15. NUMBER OF PAGES 20	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL	

## CONTENTS

	PAGE
SUMMARY.....	1
INTRODUCTION.....	1
METHOD.....	3
Subjects.....	3
Measures.....	3
Procedures.....	6
RESULTS.....	6
DISCUSSION.....	7
REFERENCES.....	10

## TABLES

Table No.	
1	AFOQT Form O Tests and Composites..... 4
2	Descriptive Statistics for UNT and UPT Criteria..... 5
3	Regression Results for the Ten Criteria..... 6

## PREFACE

The research reported in this paper was a thesis effort completed by the first author and supervised by the second. The authors would like to thank Jacobina Skinner, Linda Sawin, Melody Darby, and Stephen Larsen for their expertise with aircrew data and statistical analyses. Additionally, Bill Wimpee, Mark Teachout, and "The CAC" are thanked for their support of this study.

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/Availability	
Aviation Research Center	
Distribution/Availability	
By	
Distribution/Availability	

A-1

## PREDICTING AIRCREW TRAINING PERFORMANCE WITH PSYCHOMETRIC $g$

### SUMMARY

A comparison of the validity of general cognitive ability,  $g$ , and specific ability,  $s$ , for predicting pilot and navigator criteria was conducted. General cognitive ability and specific abilities were derived from a multiple aptitude test battery. The criteria included academic performance, and ratings of flying maneuvers such as landings, loops and rolls for pilots, and airborne navigation tasks such as day and night celestial fixes and locations for navigators. Regression analyses were conducted to evaluate the predictive efficiency of  $g$  and  $s$ . Despite the wide variability of the appearance of the criteria,  $g$  was the best predictor of all criteria and  $s$  contributed little beyond  $g$ . The average validity for  $g$  across all pilot and navigator criteria was .332 while the average validity for the specific abilities was .068. The incremental validity of specific abilities beyond the prediction afforded by  $g$  for pilot and navigator criteria, averaged .08 and .02 respectively. Results suggested that the incremental validity of specific measures for pilots may be due to specific knowledge about aviation principles and aviation instruments and aircraft controls. No navigator specific knowledge items were available in the test.

### INTRODUCTION

Although general cognitive ability was first proposed by Galton, it was early in the 20th century when Charles Spearman (1904) noted the positive correlations among mental ability tests of various content; a phenomenon termed positive manifold and a direct consequence of general cognitive ability. Encouraged by his mentor, Karl Pearson, Spearman developed the statistical technique of factor analysis through which he identified the factor responsible for the tests' correlations (Aiken, 1982). This factor he labeled  $g$ , for general factor or general ability. In addition to  $g$ , his original model included  $s_1, s_2, s_3, \dots, s_n$ , representative of specific factors unique to each test. These specific factors would not be shared among tests, unlike group factors which might be common to two or more tests but which were not correlated with  $g$ .

Psychometric  $g$  typically accounts for the majority of the test variance and usually exceeds the variance accounted for by all of the specific abilities combined (Jensen, 1980). Some (Humphreys, 1989) claim that  $g$  is unstable as it varies depending on the statistical estimation method. However, Ree and Earles (1991a) and Earles and Ree (1991) showed that unrotated principal components, unrotated principal factors, and hierarchical factor analysis estimated  $g$  with little difference so long as sufficient positive manifold existed. Their  $g$  estimate correlations ranged from .930 to .999 with most above .990.

As American psychologists investigated mental ability, they shifted from  $g$  and Spearman's Two Factor theory to the notion that cognitive ability was composed of many and varied specific abilities. This is often called the theory of differential ability, the specificity doctrine (Jensen, 1984) or the multifactor theory. Among the multifactor theorists were E. L. Thorndike, C. Hull, and L. L. Thurstone. Thorndike (1927) proposed a model that consisted of

social, concrete, and abstract intelligence. Hull (1928) developed the concept of substitutability of specific skills for general ability, however he did not provide empirical evidence for this work.

Thurstone (1938) in publishing his very influential test, Primary Mental Abilities, originally denied a correlation among his primary mental abilities (factors) that accounted for intelligence. However, he eventually acknowledged that the factors were correlated and that  $g$  was required to account for the correlations (Thurstone & Thurstone, 1941; see also Holzinger & Harman 1938; Spearman, 1939). Despite the evidence against differential ability theory, American psychologist continued seeking multiple abilities by means of tests which differed in appearance, an example of the Topographic Fallacy (Walters, Miller, & Ree, in press).

Unlike American psychologists, British psychologists, especially Philip Vernon, persisted in the investigation of  $g$ . Vernon (1960) proposed a hierarchical model of intelligence which was related to Spearman's theory. Two major group factors, (as opposed to the specific factors of Spearman) "verbal education" and "practical-mechanical-spatial," composed of  $g$  and specific abilities, occupied lower levels in his hierarchical model. Though Vernon's work was empirically sound, its impact on American psychology was small and most research continued to focus on multifactor theories.

However, empirical evidence for the predictive efficacy of  $g$  continued to accumulate. For example, McNemar (1964) reported that multiple aptitude batteries achieved little differential validity (Brogden, 1951) compared to tests designed to measure general ability. He reviewed 4,096 validity coefficients of one such test, the Differential Aptitude Test (Bennett, Seashore, & Wesman, 1982), and reported that only four of the eight subtests demonstrated "adequate" differential validity. Two of the four subtests, Verbal Reasoning and Numerical Ability, were very similar to the content of intelligence tests and provided good estimates of general ability.

Recent empirical studies have again shown the value of  $g$  as a predictor of practical criteria (Carey, 1992; Hunter & Hunter, 1984; McHenry, Hough, Toquam, Hanson, & Ashworth, 1990; Ree & Earles, 1991b, 1992; Ree, Earles, & Teachout, 1991; Thorndike, 1985, 1986). When training criteria were regressed on general and specific abilities,  $g$  was more predictive than specific abilities. Hunter and Hunter (1984) summarized the results of 515 General Aptitude Test Battery (GATB) validity studies performed over 35 years. Validity coefficients for  $g$  varied across five job families grouped by level of job complexity. They ranged from .49 to .59 with an average of .53 and were likely underestimated because they were not corrected for range restriction.

Results from the Army's Project A showed the same results for job performance criteria. McHenry, Hough, Toquam, Hanson, and Ashworth (1990) found that  $g$  was the best predictor of job performance and that adding specific ability measures increased prediction (incremental validity) by .02 or less.

More recently, Ree and Earles (1991b) regressed 78,041 airmen's technical school grades on  $g$  and  $s_1, \dots, s_n$  estimated from a multiple aptitude test battery. For all 82 jobs examined,  $g$  was the most valid predictor with the non- $g$  portions

of the test yielding an average increase in predictiveness of about .02, much like the result found by McHenry et al. (1990) and Hunter and Hunter (1984).

Ree, Earles and Teachout (1991) conducted a similar study using job performance criterion measures and found similar results;  $g$  was the best predictor and specific measures incremented predictive validity .06. Carey (1992) also conducted a study using job performance as a criterion and found increments above  $g$  of about .02. Again these results were similar to those of McHenry et al. (1990), and Hunter and Hunter (1984).

Selection is becoming increasingly important in the face of fewer military training resources and expected increases in job complexity. Despite the empirical evidence of  $g$ 's superior predictive validity for training and performance criteria, the Air Force uses measures from a multiple aptitude battery, claimed to be specific for the prediction of pilot and navigator success. If  $g$  were a better predictor of the criteria than specific abilities, selection agencies would be better off with a composite which was highly  $g$  loaded rather than with a highly specific composite.

The purpose of the current study was to investigate the contribution of  $g$  and  $s$  to the prediction of pilot and navigator criteria.

## METHOD

### Subjects

The subjects were approximately 1,400 Undergraduate Navigator Training (UNT) students and 4,000 Undergraduate Pilot Training (UPT) students who tested on Form O of the Air Force Officer Qualifying Test (AFOQT) between 1981 and 1985.

At the time of testing a majority of the subjects possessed a high school education and more than 50 percent had obtained some college education. All had baccalaureate degrees when they began pilot or navigator training. The sample included subjects commissioned through the Air Force Reserve Officer Training Corps or Officer Training School. The sample did not include Air Force Academy graduates as they do not take the AFOQT.

### Measures

As shown in Table 1, the AFOQT is composed of sixteen tests, three of which are classified as power tests: Mechanical Comprehension, Rotated Blocks, and General Science. Electrical Maze, Instrument Comprehension, and Block Counting are primarily speeded and the remaining tests are of a mixed power and speed model (Skinner & Ree, 1987). The tests are assembled into five composites used for officer selection and classification of pilots and navigators: Verbal (V), Quantitative (Q), Academic Aptitude (AA), Pilot (P), and Navigator-Technical (N-T). These composites are a reification of the belief in differential aptitude theory, however, they are all highly  $g$  saturated (Earles & Ree, 1991).

The predictors were the sixteen principal components extracted from the AFOQT. All scores were from first-time administration to avoid practice effects.



Principal components analysis (Hotelling, 1933a, 1933b) yields orthogonal components, the first of which represents the majority of variance in the data,  $g$ . The number of components extracted was equal to the number of tests. The first principal component extracted from an aptitude battery is typically a measure of  $g$ . The remaining components represent specific ability measures ( $s_1, \dots, s_n$ ). Rotation was not performed, because it redistributes first factor variance among the remaining factors. Rotation would mean that the first factor is no longer an adequate measure of  $g$  and that all the factors measure  $g$  to some extent.

Note that all of the factorial variance including the variance of the group factors and the specific variance of each test is included in the set of unrotated principal components. Scores for each principal component were calculated for each subject from weights estimated by Earles and Ree (1991).

Five UNT and five UPT grades or ratings of work samples were the criteria. Two criteria were dichotomous and eight were continuous. The dichotomous variables were the UNT and UPT Pass-Fail Final School Grades. A "pass" was reported if the overall grade average exceeded 70. Eighty-four percent of the UNT subjects and 79 percent of the UPT subjects passed training.

Table 1. AFOQT Form O Tests and Composites

Subtests	Items	Time	Composites <sup>a</sup>				
			P	N-T	AA	V	Q
Verbal Analogies	25	8	X		X	X	
Arithmetic Reasoning	25	29		X	X		X
Reading Comprehension	25	18			X	X	
Data Interpretation	25	24		X	X		X
Word Knowledge	25	5			X	X	
Math Knowledge	25	22		X	X		X
Mechanical Comprehension	20	22	X	X			
Electrical Maze	20	10	X	X			
Scale Reading	40	15	X	X			
Instrument Comprehension	20	6	X				
Block Counting	20	3	X	X			
Table Reading	40	7	X	X			
Aviation Information	20	8	X				
Rotated Blocks	15	13		X			
General Science	20	10		X			
Hidden Figures	15	8		X			
Total	350	208					

a. P is Pilot composite, N-T is Navigator-technical composite, AA is Academic Aptitude composite, V is Verbal composite, and Q is Quantitative composite.

In addition to the dichotomous pass-fail criterion, there were four other ratings-based UNT criteria. They included Airmanship Grade, Basic Procedures Grade, Day Celestial Check Flight Rating, and Night Celestial Check Flight

Rating. The content of the Airmanship course section included instruction on flight instruments and mapreading. The Basic Procedures course included flight safety, airspace, and earth physics training. Day Celestial Check Flight and Night Celestial Check Flight Ratings were work sample measures of stellar observations, sun plotting, and actual flight missions. The grades and ratings could range from 0 to 100.

UPT criteria included pass-fail, Phase 2 Check Ride average, Phase 3 Check Ride average, Air Training Command (ATC) Phase 2 Average, and ATC Phase 3 Average. Check ride averages (work samples) were ratings of actual flight missions flown in jet aircraft, the fighter-like T-37 and T-38. Phase 2 involved initial jet training in the T-37 and Phase 3 consisted of advanced flight instruction in a sophisticated supersonic aircraft, the T-38. Phase averages were cumulative grades covering flying performance, commanders' ratings, and written tests on various subjects such as mission planning and other aspects of airmanship. UPT Phase 2 and 3 ratings and course grades could range from 0 to 100. See Table 2 for descriptive statistics of the UNT and UPT criteria.

All work sample ratings were made by instructor pilots or by instructor navigators. These ratings are routinely collected as part of their duties. No reliability estimates were available for the criteria.

Table 2. Descriptive Statistics for UNT and UPT Criteria

Criteria	N	Max	Min	Mean	SD
UNT					
Pass/Fail	1411	1.00	0.00	0.84	0.36
Airmanship	1341	100.00	60.00	93.59	6.00
Basic Procedures	1176	100.00	50.90	93.23	6.54
Day Check Flight	1224	100.00	0.00	87.80	13.33
Night Check Flight	1182	100.00	0.00	85.60	15.40
UPT					
Pass/Fail	3942	1.00	0.00	0.79	0.40
Phase 2 Check Ride	2203	98.90	6.42	84.69	15.23
Phase 3 Check Ride	1867	100.00	21.00	90.52	8.08
Phase 2 Average	2203	92.14	6.54	72.04	13.02
Phase 3 Average	1867	93.73	24.34	81.59	7.40

Training success is often considered a more vital criterion than job performance because it is an antecedent. It is more advantageous to detect a poor performer prior to or during training rather than afterwards. Training grades have been utilized in research by many including Hunter (1986), Hunter and Hunter (1984), Schmidt and Hunter (1978), Arth, Steuck, Sorrentino, and Burke (1989), and Ree and Earles (1991b).

## Procedures

A total of ten stepwise multiple regressions were computed on the raw data. An analogous set of regressions was run after the data were corrected for range restriction (Lawley, 1943), however, the variables included in the regressions were only those which were found to be significant in the regressions computed in the data prior to correction for range restriction. No statistical tests were conducted in the data after correction for range restriction. The Type I error rate was set at  $p < .01$ .

Because the range restriction correction increases sampling error variance of corrected correlations, effective sample size estimate were used in the cross validation procedures. Using the original sample size in the estimates of cross validated correlations would bias the estimates upward. Schmidt, Hunter, and Larson (1988) noted that the increase in standard error of corrected correlations was equivalent to using a smaller sample size and solved the usual standard error of  $r$  for this effective sample size. They found that effective sample sizes were notably smaller than the original sample sizes. Multiple correlation coefficients along with the effective sample sizes were then used in the computation of the Stein's expectancy operator (Kennedy, 1982) to estimate the reduction in the multiple correlation coefficients that would occur on cross validation.

## RESULTS

The five UNT and five UPT criteria were predicted with samples ranging from 1,176 to 3,942 subjects. Table 3 shows the results of the regression analyses.

Table 3. Regression Results for the Ten Criteria

Criterion	Uncorrected <i>rg Rg+s</i>		Corrected <i>rg Rg+s</i>		Cross-Validated <i>Rcg+s Diff</i>	
UNT						
Pass-Fail	.248	.311	.375	.429	.409	.034
Airmanship	.372	.406	.509	.532	.515	.006
Basic Procedures	.366	.390	.523	.556	.536	.013
Day Check Flight	.136	.172	.242	.292	.290	.048
Night Check Flight	.159	.228	.254	.313	.279	.024
UPT						
Pass-Fail	.170	.304	.284	.376	.366	.082
Phase 2 Check Ride	.204	.361	.338	.445	.431	.093
Phase 3 Check Ride	.131	.210	.209	.283	.263	.053
Phase 2 Average	.211	.390	.352	.467	.455	.102
Phase 3 Average	.141	.232	.237	.312	.295	.058

$Rcg+s$  is the corrected cross validated correlation using the Stein Estimator with the effective sample size. Phase 2 and 3 are cumulative averages.

The column headed  $rg$  is the bivariate correlation denoting the predictive efficiency of  $g$ , and  $Rg+s$  is the multiple correlation of  $g$  and  $s_1 \dots s_{15}$  with the criteria. The  $Rg+s$  values reflect all the measures that entered the regression equations. The differences between  $rg$  and  $Rg+s$  which indicate the strength of specific abilities as predictors appear in the column labeled "Diff."

General ability entered first in all (uncorrected and corrected) but two uncorrected regression equations. For the UPT Phase 3 Check Ride Average and UPT ATC Phase 3 Average as criteria, the second principal component ( $s_1$ ) entered first in the uncorrected regressions. This phenomenon may reasonably be attributed to the artifactual distortions caused by the effects of prior selection on the uncorrected correlation matrices. Aside from  $g$ , only 4 specific measures entered frequently while seven of the specific measures never entered. In other words, seven specific measures added nothing to prediction.

The cross validation estimates of the multiple correlation coefficients were computed with the Stein Estimator (Kennedy, 1982). Effective sample sizes constituted part of the calculation. In both the navigator and pilot groups cross validation brought an average reduction in multiple correlation of approximately .015.

## DISCUSSION

The data clearly demonstrated that  $g$  was the best predictor for all the criteria. Corrected  $rgs$  ranged from .209 to .523; corrected  $rss$  ranged from .023 to .115. General ability's average validity coefficient was .332 versus the average of specific abilities of .068 and there was no overlap in the two ranges.

Specific abilities contributed a little to the prediction of the criteria. The average increment to validity due to specific abilities across the five navigator criteria was .02 and across the five pilot criteria was .08. The smallest increment by specific ability (.006) to the validity of  $g$  was for navigator Airmanship, a job knowledge criterion with aerodynamics, flight instrument and cockpit familiarization, and aircraft emergency procedure content. This is consistent with the belief that  $g$  is strongly related to learning ability (Jensen, 1986). The largest increment to  $g$  (.102) was for the pilot Phase 2 Average. Overall, specific abilities exhibited greater incremental validity for the pilot criteria than for the navigator criteria.

Those specific abilities which were predictive of navigator criteria did not overlap with the specific abilities which were predictive of pilot criteria. Specific abilities predictive of navigator criteria were not consistent across all navigator criteria. Only  $g$  was found in every navigator prediction. There was little in common among equations. For the pilot, three predictors entered every equation:  $g$ ,  $s_1$ , and  $s_3$ . Although the psychological nature of  $s_1$  and  $s_3$  can not be assessed with any certainty, they emphasized special knowledge of aviation information and instrument comprehension. This special knowledge appears to be an example of Cattell's (1987) crystallized intelligence. Cattell's theory includes both a fluid intelligence which is available to learn anything and crystallized intelligence(s) which is the product of learning. Crystallized intelligence refers to knowledge or skills acquired by the "investment" of "fluid" intelligence in learning some information such a specialized knowledge

of flying. For example, Carretta and Ree (in press) found that specialized knowledge of aircraft instruments, controls and aviation terms as manifested by the number of hours flown prior to entering pilot training was a good predictor of pilot training performance.

The current finding of the predictiveness of  $s_1$  and  $s_2$  is consistent with the predictiveness findings of Carretta and Ree (in press). The test used to measure  $g$  and  $s$  did not have subtests which measured specialized knowledge about navigation. There were no questions about sextants, star transits, or global positioning system satellites. Had tests measuring navigator specific knowledge been available, greater effects might have been found for specific ability or knowledge for navigators. However, it is not clear that applicants are exposed to such information as frequently as to pilot and aircraft information. This would almost certainly cause the validity of these navigator special knowledge tests to be low. The use of specific knowledge tests may pose this kind of problem for many, if not most, jobs. Further studies of the incremental validity of specific knowledge or crystallized intelligence should be accomplished to illuminate the issue.

The policy consequences of using this specific knowledge or crystallized intelligence as a predictor should also be investigated, especially for women and members of minority groups who are less likely to be exposed to information about flying and navigation.

Additionally, the increment found for pilots in this study was equal to the increment found in Carretta and Ree (in press) who used several different measures of specific ability. A meta analysis could clarify the relationship of these two findings.

However, like most correlations, those presented here should not be interpreted at face value. It should be noted that these average incremental validity values are probably upwardly biased. This is because the correlation of  $g$  and the criteria is a bivariate correlation which is a downwardly biased estimator and the correlation of  $g+s$  with the criteria is a multiple correlation which is an upwardly biased estimator. The true difference between them is therefore less than shown.

Differences in criterion reliability and absolute level of criterion reliability effect validity correlations (Hunter, Schmidt, & Jackson, 1982). The magnitude of a correlation is dependent on reliability of the variables involved. Criterion reliabilities are likely not all the same and would therefore have increased the observed variability of both the correlations of  $g$  with the criteria and the specific abilities with the criteria. As no estimates of criterion reliabilities were available, no corrections could be made.

Overall,  $g$  was more predictive of navigator than pilot criteria. The corrected correlations of  $g$  with the navigator criteria ranged from .242 (UNT Day Celestial Check Flight) to .523 (UNT Basic Procedures) with a mean of .380. With the UPT criteria the correlations ranged from .209 (UPT Phase 3 Check ride) to .352 (UPT Phase 2 Average) with a mean of .284. This difference in average correlational magnitude may be due to course content differences or to differences in reliability of the criterion measures. The cause can not be known

from these data.

Additionally the corrected correlation coefficients were also likely underestimates of the relationship between the  $g$  and  $s_1 \dots s_{15}$  and the criteria because the correction was back to a group of applicants who were stringently selected for college entry. This is consistent with Thorndike's (1986) explanation:

One reason that a measure of cognitive ability sometimes does not show up so favorably in relation to other more specialized tests, or in relation to noncognitive measures, is that prior test, educational, or life hurdles have already screened out those low in  $g$ , who would have been likely to fail because of limits of cognitive ability. (p.338).

That the 2 pass-fail criteria are dichotomous and have low variability may also contribute to underestimation. The observed coefficients may not be far from the maximum obtainable observed correlations.

Three artifacts need to be considered in interpreting the results. No criterion reliability was available. The correlations were not completely corrected for prior selection and the dichotomous criteria had extreme splits. These three artifacts hampered our understanding of the results or reduced the observed correlations.

These results extend the findings for  $g$  beyond previous research to new samples. They confirm the value of  $g$  as a predictor of additional criteria. Again, the incremental validity of  $s_1 \dots s_n$  was small. Taken together with previous results, general cognitive ability continues to appear as the universal predictor of job and training performance. From jelly rolls (Jensen, 1980) to aileron rolls,  $g$  predicts criteria of interest.

## REFERENCES

- Aiken, L. R. (1982). *Psychological testing and assessment*. Boston, MA: Allyn and Bacon, Inc.
- Arth T. O., Steuck, K. W., Sorrentino, C. T., & Burke, E. F. (1989). *Air Force Officer Qualifying Test (AFOQT): Predictors of undergraduate pilot training and undergraduate navigator training Success* (AFHRL-TP-89-52). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.
- Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1982). *Differential Aptitude Tests (Forms V and W): Administrators handbook*. New York: Psychological Corporation.
- Brogden, H. E. (1951). Increased efficiency of selection resulting from replacement of a single predictor with several differential predictors. *Educational and Psychological Measurement*, 11, 173-196.
- Carey, N. (1992). *New predictors of mechanics' job performance: Marine Corps findings*. A paper presented at the annual meeting of the American Psychological Association, Washington, D. C.
- Carretta, T. R., & Ree, M. J. (in press). Pilot candidate selection method (PCSM): What makes it work?
- Cattell, R., (1987). *Intelligence, its structure, growth and action*. New York: Elsevier.
- Earles, J. A., & Ree, M. J. (1991). *Air Force Officer Qualifying Test (AFOQT): Estimating the general ability component* (AL-TP-1991-0039). Brooks AFB TX: Manpower and Personnel Division, Human Resources Directorate, Armstrong Laboratory.
- Holzinger, K. J., & Harman, H. H. (1938). Comparison of two factorial analyses. *Psychometrika*, 3, 45-60.
- Hotelling, H. H. (1933a). Analysis of a complex of statistical variables with principal components. *Journal of Educational Psychology*, 24, 417-441.
- Hotelling, H. H. (1933b). Analysis of a complex of statistical variables with principal components (continued). *Journal of Educational Psychology*, 24, 498-520.
- Hull, C. L. (1928). *Aptitude testing*. Yonkers, N.Y.: World Book Co.

- Humphreys, L. G. (1989). The first extracted factor is an unreliable estimate of Spearman's "g": The case of discrimination reaction time. *Intelligence*, 13, 319-323.
- Hunter, J. E. (1986). Cognitive ability, cognitive aptitude, job knowledge, and job performance. *Journal of Vocational Behavior*, 29, 340-362.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternate predictors of job performance. *Psychological Bulletin*, 96, 1, 72-98.
- Hunter, J. E., Schmidt, F. L., & Jackson, G. (1982). *Meta analysis; Cumulating results across studies*. Beverly Hills, CA: Sage.
- Jensen, A. R. (1980). *Bias in mental testing*. The New York: Free Press.
- Jensen, A. R. (1984). Test validity: g versus the specificity doctrine. *Journal of Social and Biological Structures*, 7, 93-118.
- Jensen, A. R. (1986). g: artifact or reality? *Journal of Vocational Behavior*, 29, 301-331.
- Kennedy, E. (1982). Estimation of the squared cross-validation coefficient in the context of best subset regression. *Applied Psychological Measurement*, 231-237.
- Lawley, D. N. (1943). A note on Karl Pearson's selection formulas. *Proceedings of the Royal Society of Edinburgh, Section A*, 62, Part I, 20-30.
- McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology*, 43, 335-354.
- McNemar, Q. (1964). Lost: our intelligence? Why? *American Psychologist*, 19, 871-882.
- Ree, M. J., & Earles, J. A. (1991a). The stability of convergent estimates of g. *Intelligence*, 15, 271-278.
- Ree, M. J., & Earles, J. A. (1991b). Predicting training success: Not much more than g. *Personnel Psychology*, 44, 321-332.
- Ree, M. J., & Earles, J.A. (1992). Intelligence is the best predictor of job performance. *Current directions in psychological science*, 1, 86-89.



- Ree, M. J., Earles, J.A., & Teachout, M. (1991). *General cognitive ability predicts job performance*. (AL-TP-1991-0057), Manpower and Personnel Research Division, Human Resources Directorate, Armstrong Laboratory, Brooks AFB: Texas.
- Schmidt, F. L., & Hunter, J. E. (1978). Moderator research and the law of small numbers. *Personnel Psychology*, 31, 215-232.
- Schmidt, F. L., Hunter, J. E., & Larson, M. (1988). General cognitive ability vs. general and specific aptitudes in the prediction of training performance: Some preliminary findings. In W. H. Angoff (Chair), *Promising areas for psychometric Research*. Symposium conducted at the meeting of the American Psychological Association Convention, Georgia, August.
- Skinner, J., & Ree, M. J. (1987). *AFOQT Item and Factor Analysis of Form O* (AFHRL-TR-86-68). Brooks AFB TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.
- Spearman, C. (1904). "General Intelligence" objectively determined and measured. *American Journal of Psychology*, 15, 210-293.
- Spearman, C. (1939). Thurstone's work reworked. *Journal of Educational Psychology*, 39, 1-16.
- Thorndike, E. L. (1927). *Measurement of intelligence*. New York: Teacher's College, Columbia University.
- Thorndike, R. L. (1985). The central role of general ability in prediction. *Multivariate Behavioral Research*, 20, 241-254.
- Thorndike, R. L. (1986). The role of general ability in prediction. *Journal of Vocational Behavior*, 29, 332-339.
- Thurstone, L. L. (1938). *Primary mental abilities*. *Psychometric Monographs*, No. 1.
- Thurstone, L. L., & Thurstone, T. G. (1941). Factorial studies of intelligence. *Psychometric Monographs*, No. 2.
- Vernon, P. E. (1960). *Intelligence and cultural environment*. London: Methuen.
- Walters, L. C., Miller, M., & Ree, M. J. (in press) Structured interviews for pilot selection: No incremental validity.